REMOTE ACCESS TOOL FOR EARTH SCIENCE DATA

Elaine Dobinson and Robert Raskin*
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
Elaine. Dobinson @jpl.nasa.gov and Robert. Raskin@jpl.nasa.gov

Abstract

This demo presents an http-based client/server application prototype that facilitates internet access to Earth science data. The client consists of a Java applet GUI that allows the user to select spatial/temporal subsets of indexed datasets. The client also includes a MA TLAB interface that allows the incoming data to be loaded directly into a MATLAB session. The server provides directory, catalog, and data access services and performs the subsetting operations prior to data transmission. An example is presented where data from multiple sources and in multiple formats are combined into a single MATLAB plot. The prototype addresses the lack of common data models in the Earth sciences. It also addresses the needfor access to corroborating data by Earth Observation Satellite (EOS) instrument team members for calibration and validation.

The Earth Science Remote Access Tool builds on the Distributed Oceanographic Data System (DODS) developed at the University of Rhode Island and the Massachusetts Institute of Technology [1, 2]. DODS provides a common data format with translators for many standard formats (netCDF, MATLAB, DSP or JGOFS) and data models (array, swath, grid, etc.). A DODS server translates from the data models of these formats into the intermediate DODS model. A DODS client performs the reverse translation, such as into MATLAB or netCDF. As a result, local applications expecting local data in a particular format can access remote datasets in other formats. A DODS/HDF server was built as part of this effort and supports the SDS, raster, and Vdata elements of the HDF data model.

---

* The work reported in this paper is being performed at the Jet Propulsion Laboratory of the California Institute of Technology under contract with the National Aeronautics and Space Administration.

The DODS data model includes support for "Arrays", "Grids", and "Sequences." For Arrays and Grids, support is provided for subletting by row and column. For this prototype, coverage indexes for several oceanographic datasets were developed to map the rows/columns to spatial/temporal coordinates. This permits DODS clients to query by geographical and temporal coordinates and receive only the desired subsets and variables.

The system architecture is outlined in Figure 1. Searches are carried out by accessing directory and catalog servers, which are implemented as C++ CGI programs. The master directory contains a list of dataset holdings at local or remote sites and their associated URLs. The dataset catalogs contain the spatial and temporal bounds for elements of each dataset. For swath data, each cross-track is indexed. When a user requests a spatial/temporal subset of swath data, the *server* concatenates any contiguous cross-tracks lying within the bounding rectangle and time range. For gridded data, each grid is indexed; a user requesting a subset receives back the subgrid corresponding to their spatial/temporal selections.

At the start of a typical session, the user is provided with a list of datasets residing in the master catalog. The user selects the datasets of interest as well as a spatial region and time range, using the spatial/temporal selector tool. The server responds with a list of data subsets that satisfy the constraints and presents a list of the variables available. After the user selects the desired subsets and variables, the subset requests are *converted* to DODS commands to access the requested data. The data server need not reside at the same location as the catalog and directory servers. The MATLAB client reads the data by invoking a helper application in the client web browser.

The prototype currently is tailored toward a particular application: the calibration and validation of the NASA scatterometer (NSCAT). This EOS instrument measures winds over the global oceans and began data collection last year. [n order to determine the accuracy of the instrument, NSCAT scientists need access to many different ground-truth and satellite datasets coincident in space and time with the NSCAT data. Several of these datasets are included in the master catalog.

One of the lessons learned in the development of the prototype was that 100% interoperability between formats is non-trivial. HDF data structures are richer than those in netCDF or in the intermediate DODS format. Some structural information may be lost in the conversion. In particular, our DODS server currently does not provide translation support for HDF Vgroups. However, future plans include such enhancements.

This work was sponsored by the ESDIS Prototyping Office and is ongoing. Future plans include graphical display of database coverages in space and time. Other plans are to provide support for the HDF-EOS format, as HDF-EOS files contain the spatial and temporal bounds as embedded metadata. In addition, it would be desirable to redesign the system to utilize Java servlets and to provide support for indexing on dependent variables, to aid content-based searches. A White Paper describing the project in greater detail can be found at: http://dods.jpl.nasa.gov/wp. Software support was provided by Todd Karakashian, Isaac Henry, and David Hecox. Elaine Dobinson, Deputy Task Manager for the Physical Oceanography Distributed Archive Archive Center (PO. DAAC) at JPL, is the task leader, and Robert Raskin is the current developer.

References

[1] Gallagher, J. and G. Milkowski, "Data Transport Within the Distributed Oceanographic Data System", *4th **International World Wide** Web Conference Proceedings, 691-700, 1995.*

[2] Sgouros, T., ***DODS** User Guide, Version 1.0, 1996.*

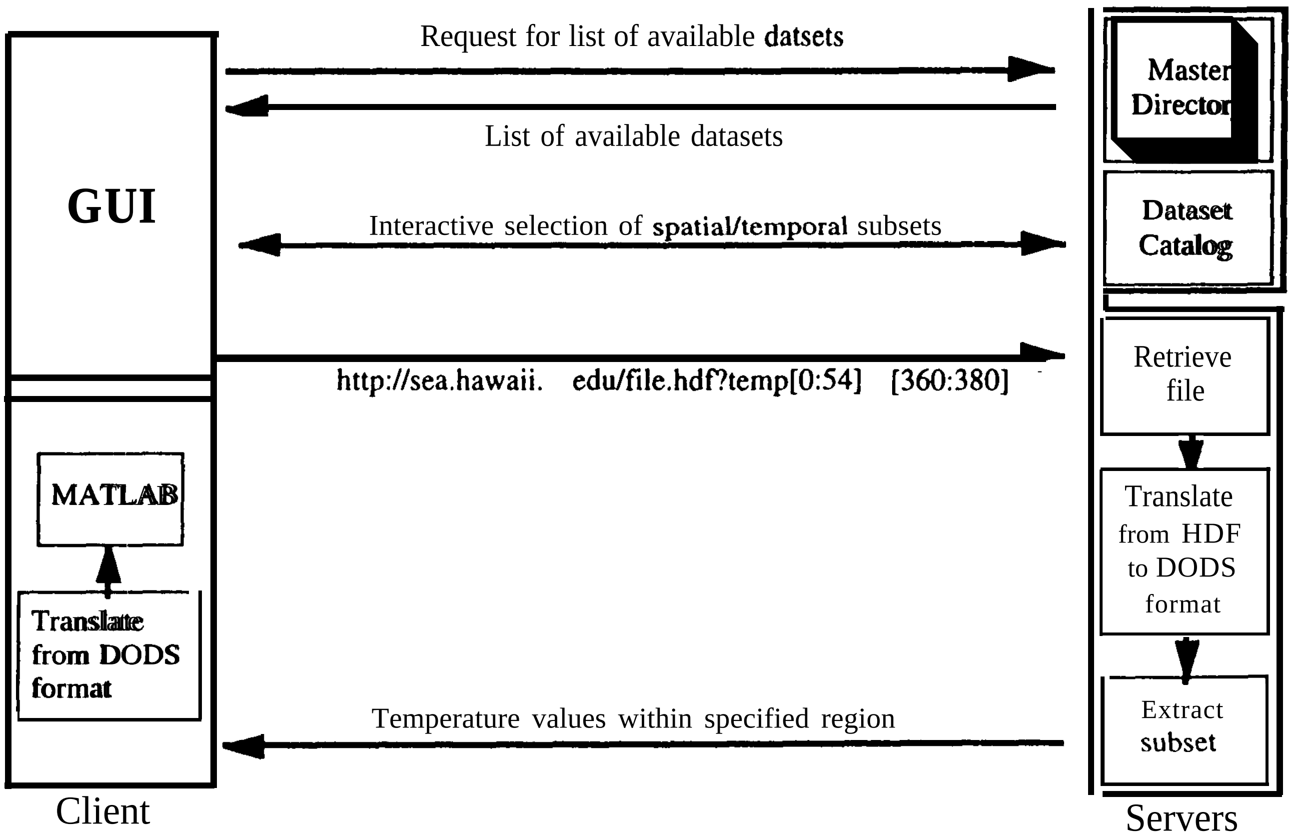

Figure 1. System architecture